Communications of the Korean Statistical Society
2009, Vol. 16, No. 4, 713–722

Effective Computation for Odds Ratio Estimation in Nonparametric Logistic Regression

Young-Ju Kim[1,a]

[a]Department of Information Statistics, Kangwon National University

Abstract

The estimation of odds ratio and corresponding confidence intervals for case-control data have been done by traditional generalized linear models which assumed that the logarithm of odds ratio is linearly related to risk factors. We adapt a lower-dimensional approximation of Gu and Kim (2002) to provide a faster computation in nonparametric method for the estimation of odds ratio by allowing flexibility of the estimating function and its Bayesian confidence interval under the Bayes model for the lower-dimensional approximations. Simulation studies showed that taking larger samples with the lower-dimensional approximations help to improve the smoothing spline estimates of odds ratio in this settings. The proposed method can be used to analyze case-control data in medical studies.

Keywords: Bayesian confidence interval, case-control, odds ratio, smoothing splines.

1. Introduction

A logistic regression model with binary response data is

$$\log\left(\frac{p(\mathbf{x})}{1-p(\mathbf{x})}\right) = \eta(\mathbf{x}),$$

where $p(\mathbf{x}) = P(Y = 1|\mathbf{x})$, Y is the binary response taking values 0 or 1 and $\mathbf{x}$ is a vector of covariates. A typical generalized linear model assumes $\eta(\mathbf{x}) = \mathbf{x}^T\beta$. However, the parametric assumptions may be too rigid in some applications. Semiparametric or nonparametric function estimation techniques have been developed by many researchers so that they can relax such a strong assumptions in a certain function form. Among various nonparametric function estimation methods, smoothing splines have been popularly used in nonparametric regression settings. see Gu (2002) for details. However, it suffers practical limits due to heavy computation for large data. In order to overcome such computational burden, various approximations were suggested (see Luo and Wahba, 1997; Xiang and Wahba, 1998; Lin *et al.*, 2000; Gu and Kim, 2002). Gu and Kim (2002) suggested an effective lower-dimensional approximations for faster computation of smoothing splines for exponential families by using a random subset of kernel basis. Kim (2003) and Kim and Gu (2004) suggested a Bayes model for the lower-dimensional approximations in Gaussian smoothing splines. Wang (1997) considered the estimation of odds ratio by using smoothing splines technique and showed that the Bayesian confidence interval for the odds ratio and its bias-corrected odds ratio with full basis has Wahba (1983)'s frequentist across-the-function coverage properties. This paper extends Kim and Gu (2004)'s Bayes model for the lower-dimensional approximations to nonparametric logistic regression where the estimation of the odds ratio in smoothing spline settings and its Bayesian confidence intervals are of

[1] Assistant professor, Department of Information Statistics, Kangwon National University, Chuncheon 200-090, Republic of Korea. E-mail: ykim7stat@kangwon.ac.kr

primary interest. We explore whether the Bayesian confidence interval for the odds ratio with the lower-dimensional approximation has similar frequentist properties through simulation studies.

The rest of the article is organized as follows. In Section 2, smoothing splines and Bayesian model based on the lower-dimensional approximation are reviewed. In Section 3, odds ratio estimation and its Bayesian confidence interval based on the lower-dimensional approximation are proposed. Section 4 presents the computation of Bayesian confidence intervals. Monte Carlo simulation results are shown in Section 5.

2. Smoothing Splines and Bayes Model

Assume that Y_i is i^{th} binary response, $i = 1, \ldots, n$, and x_i is a covariate of i^{th} subject. A smoothing spline in nonparametric logistic regression is the minimizer of the following penalized logistic likelihood functional

$$-\frac{1}{\sigma^2}\sum_{i=1}^{n}\{Y_i\eta(x_i) - \log(1+\exp(\eta(x_i)))\} + \frac{n\lambda}{2}J(\eta), \tag{2.1}$$

where $\eta(x_i)$ is unknown smooth function, smoothing parameter λ plays the important role of controlling the trade-off between goodness-of-fit and roughness, and $J(\eta)$ is a penalty function. In fact, the minimizer of (2.1) lies in $\mathcal{H} = \mathcal{N}_J \oplus \mathcal{H}_J$, where $\mathcal{N}_J = \{\eta : J(\eta) = 0\}$ is the null space of $J(\eta)$, the space $\mathcal{H}_J$ is a reproducing kernel Hilbert space(RKHS) with $J(\eta)$ as the square norm, and $\oplus$ indicates tensor sum decomposition. Note that a space $\mathcal{H}$ in which the evaluation functional $[x]f = f(x)$ is continuous is called a RKHS possessing a reproducing kernel(RK) $R(\cdot,\cdot)$, a non-negative definite function satisfying $\langle R(x,\cdot), f(\cdot)\rangle = f(x)$, for all $f \in \mathcal{H}$, where $\langle\cdot,\cdot\rangle$ is the inner product in $\mathcal{H}$.

The minimizer of (2.1) can be expressed as

$$\eta(x) = \sum_{\nu=1}^{m} d_\nu\phi_\nu(x) + \sum_{i=1}^{n} c_i R_J(x_i, x), \tag{2.2}$$

where $\{\phi_\nu\}$ is a basis of $\mathcal{N}_J$ and R_J is the RK in $\mathcal{H}_J$. We may rewrite η as $\eta(x) = \eta_0(x) + \eta_1(x)$, where $\eta_0 \in \mathcal{H}_0 = \text{span}\{\phi_\nu, \nu = 1, \ldots, m\}$ and $\eta_1 \in \mathcal{H}_J$.

Gu and Kim (2002) suggested a lower-dimensional approximations for faster computation in smoothing splines to relieve the computational burden and practical limit in the use of smoothing splines. They showed that the minimizer of the penalized likelihood functional in $\mathcal{H}$ shared the same convergence rates as one in the lower-dimensional functional space $\mathcal{H}_q = \mathcal{N}_J \oplus \text{span}\{R_J(u_j,\cdot), j = 1, \ldots, q\}$, where $\{u_j, j = 1, \ldots, q\}$ are random subsets of $\{x_i, i = 1, \ldots, n\}$, as long as $q \asymp n^{2/(pr+1)+\epsilon}$, where for some $p \in [1, 2]$, $r > 1$, and $\epsilon > 0$ is arbitrary. For the cubic spline, $r = 4$ is used. Then the lower-dimensional approximation in the smoothing spline model is expressed as

$$\eta(x) = \sum_{\nu=1}^{m} d_\nu\phi_\nu(x) + \sum_{j=1}^{q} c_j R(z_j, x) = \boldsymbol{\phi}^T\mathbf{d} + \boldsymbol{\xi}^T\mathbf{c}, \tag{2.3}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\xi}$ are vectors of functions and $\mathbf{d}$ and $\mathbf{c}$ are vectors of coefficients; $q = n$ for the exact solution in $\mathcal{H}$. Substituting (2.3) into (2.1), one calculate $\mathbf{d}$ and $\mathbf{c}$ by minimizing the resulting penalized likelihood functional with respect to $\mathbf{d}$ and $\mathbf{c}$.

The Bayes model for the lower-dimensional approximations in Gaussian settings was proposed by Kim and Gu (2004). Assume that $\eta = \eta_0 + \eta_1$, where η_0 has a diffuse prior in $\mathcal{N}_J$ and η_1 has a

Gaussian process prior with zero mean and the covariance function

$$E\left[\eta_1(x_i)\eta_1(x_j)\right] = bR_J\left(x_i, \mathbf{u}^T\right)Q^+R_J\left(\mathbf{u}, x_j\right),$$

where Q^+ is the Moore-Penrose inverse of $Q = R_J(\mathbf{u}, \mathbf{u}^T)$, $\{u_j\}$ is a random subset of $\{x_i\}$. Recall the notation $\xi_j = R_J(u_j, \cdot)$ and $R^T = \xi(x^T)$. Letting $M = RQ^+R^T + n\lambda I$, where I is an identity matrix, and under the prior stated above, one has

$$E\left[\eta(x)|\mathbf{Y}\right] = \phi^T\mathbf{d} + \xi^T\mathbf{c},$$

where

$$\mathbf{d} = \left(S^TM^{-1}S\right)^{-1}S^TM^{-1}\mathbf{Y},$$
$$\mathbf{c} = Q^+R^T\left(M^{-1} - M^{-1}S\left(S^TM^{-1}S\right)^{-1}S^TM^{-1}\right)\mathbf{Y}$$

and

$$\begin{aligned}\frac{\mathrm{var}[\eta(x)|\mathbf{Y}]}{b} &= \xi^TQ^+\xi + \phi^T\left(S^TM^{-1}S\right)^{-1}\phi \\ &\quad - \phi^T\left(S^TM^{-1}S\right)^{-1}S^TM^{-1}RQ^+\xi \\ &\quad - \xi^TQ^+R^TM^{-1}S\left(S^TM^{-1}S\right)^{-1}\phi \\ &\quad - \xi^TQ^+R^T\left(M^{-1} - M^{-1}S\left(S^TM^{-1}S\right)^{-1}S^TM^{-1}\right)RQ^+\xi,\end{aligned}$$

where S is a matrix of $n \times m$ with $(i, \nu)^{th}$ row $\phi_\nu(x_i)$, R is a matrix of $n \times q$ with $(i, j)^{th}$ entry $R_J(x_i, z_j)$, and Q is a matrix of $q \times q$ with $(i, j)^{th}$ entry $R_J(u_i, u_j)$.

Wahba (1983) showed that the pointwise Bayesian confidence intervals has the frequentist coverage on average. Gu (2002) extended her across-the-function coverage property to Bernoulli data in smoothing spline with full basis. Also, Wahba *et al.* (1995) extended the results of Gu and Wahba (1993) to construct approximate Bayesian confidence intervals for the smoothing spline estimates for exponential families. We extend the Bayes model of Kim and Gu (2004) to construct the approximate Bayesian confidence intervals for odds ratio by using the lower-dimensional approximations in logistic smoothing splines model.

Theorem 1. *Suppose that η_λ is the minimizer of (2.1). Under the prior stated as above, the approximate posterior mode of η given $\mathbf{Y}$ is equal to η_λ. Furthermore, the approximate posterior distribution for η given $\mathbf{Y}$ is Gaussian with mean η_λ and variance that can be derived from penalized weighted least squares functional in Kim and Gu (2004) with $M = RQ^+R^T + n\lambda W^{-1}$, where $W = diag(w_1, \ldots, w_n)$ with $w_i = p_i(1 + p_i)$.*

3. Odds Ratio Estimation

Suppose that, for a covariate x_1, we are interested in the odds ratio of $\mathbf{x}_t = (x_{1t}, \ldots, x_n)^T$ and $\mathbf{x}_s = (x_{1s}, \ldots, x_n)^T$ when other covariates are fixed. The odds ratio of $\mathbf{x}_t$ and $\mathbf{x}_s$ is

$$\mathrm{OR}\left(\frac{\mathbf{x}_t}{\mathbf{x}_s}\right) = \frac{e^{\eta(\mathbf{x}_t)}}{e^{\eta(\mathbf{x}_s)}} = \exp\left(\eta(\mathbf{x}_t) - \eta(\mathbf{x}_s)\right).$$

In order to get the (approximate) Bayesian confidence interval for the odds ratio, we need to derive the posterior distribution of $\eta(\mathbf{x}_t) - \eta(\mathbf{x}_s)$ given $\mathbf{Y}$ in logistic regression.

Theorem 2. *Let $\rho = \tau^2/b$. Under the prior stated in Section 2, letting $\rho \to \infty$, the approximate posterior distribution of $\eta(\mathbf{x}_t) - \eta(\mathbf{x}_s)$ given $\mathbf{Y}$ is Gaussian with mean $\eta_\lambda(\mathbf{x}_t) - \eta_\lambda(\mathbf{x}_s)$ and the variance*

$$var(\eta(\mathbf{x}_t) - \eta(\mathbf{x}_s)|\mathbf{Y}) = var(\eta(\mathbf{x}_t)|\mathbf{Y}) + var(\eta(\mathbf{x}_s)|\mathbf{Y}) - 2cov(\eta(\mathbf{x}_t), \eta(\mathbf{x}_s)|\mathbf{Y}).$$

The covariance between $\eta(\mathbf{x}_t)$ and $\eta(\mathbf{x}_s)$ is

$$\begin{aligned}\frac{cov[(\eta(\mathbf{x}_t), \eta(\mathbf{x}_s))|Y]}{b} = {} & \boldsymbol{\phi}^T(\mathbf{x}_s)\left(S^TM^{-1}S\right)^{-1}\boldsymbol{\phi}(\mathbf{x}_t) - \boldsymbol{\phi}^T(\mathbf{x}_s)\left(S^TM^{-1}S\right)^{-1}S^TM^{-1}RQ^+R(\mathbf{t})^T \\ & - R(\mathbf{s})Q^+R^T\left(S^TM^{-1}S\right)^{-1}S^TM^{-1}\boldsymbol{\phi}^T(\mathbf{x}_t) + R(\mathbf{s})Q^+R(\mathbf{t})^T \\ & - R(\mathbf{s})Q^+R^T\left(M^{-1} - M^{-1}S\left(S^TM^{-1}S\right)^{-1}S^TM^{-1}\right)RQ^+R(\mathbf{t})^T,\end{aligned}$$

where $R(\mathbf{s}) = (R_J(\mathbf{x}_s, u_1), \ldots, R_J(\mathbf{x}_s, u_n))^T$.

Therefore, the approximate $100(1-\alpha)\%$ Bayesian confidence interval for the odds ratio of $\mathbf{x}_t$ and $\mathbf{x}_s$ can be constructed as

$$\left(\mathrm{OR}\left(\frac{\mathbf{x}_t}{\mathbf{x}_s}\right)\exp\left(-z_{\frac{\alpha}{2}}\sqrt{\varphi^2}\right),\ \mathrm{OR}\left(\frac{\mathbf{x}_t}{\mathbf{x}_s}\right)\exp\left(z_{\frac{\alpha}{2}}\sqrt{\varphi^2}\right)\right),$$

where $\varphi^2 = \mathrm{var}(\eta(\mathbf{x}_t) - \eta(\mathbf{x}_s)|\mathbf{Y})$.

4. Computation

The smoothing spline estimates of odds ratio can be calculated by using the smoothing spline estimates of η by minimizing (2.1) through Newton-Raphson iteration for fixed smoothing parameters. Quadratic approximations at $\eta = \tilde{\eta}$ of log penalized logistic likelihood functional lead to the penalized weighted likelihood functional with pseudo-data $\tilde{\mathbf{Y}} = \tilde{\eta} - W^{-1}\tilde{v}$, where $\tilde{v} = -\mathbf{Y} + p$ and $W = \mathrm{diag}(w_1, \ldots, w_n)$ with $w_i = p_i(1 + p_i)$. Then the resulting normal equation for $\mathbf{d}$ and $\mathbf{c}$ can be solved by Cholesky decomposition followed by forward and backward substitutions (Kim, 2003; Kim, 2009). Smoothing parameters are selected via the alternative generalized approximate cross-validation(AGACV) score of Gu and Xiang (2001). The calculation of the approximate posterior variances and covariances given in Theorem 2 can be done by taking the diagonals and off-diagonals of $\sigma A_w(\lambda)$ respectively, *e.g.*, see Kim and Gu (2004), where $A_w(\lambda)$ is the smoothing matrix. The penalty we used is $\int \ddot{\eta}$. Note that the lower-dimensional approximations in smoothing splines established in Gu and Kim (2002) and Kim and Gu (2004) were available in existing software R "gss" packages. Our methods can use it with extra modifications to calculate posterior covariances. We used $q = 10n^{2/9}$ for the lower-dimensional approximations as Kim and Gu (2004) suggested.

5. Simulation Study

Wang (1997) observed in his simulation studies that smoothing spline estimates of odds ratios are biased especially when one of two points of odds ratio is either at peak or at the valley and suggested a bootstrap bias-corrected estimate of odds ratio, but the evaluation of it was somewhat limited. The bootstrap bias-corrected estimation in smoothing spline can be done as followings. First

Table 1: Estimates of odds ratio for η_1(top) and η_2(bottom) with full basis and reduced basis(inside the parenthesis).

n	x_t(base)	x_s	True OR	$\widehat{OR}$	Stdev	95% coverage	90% coverage
100	0.5	0.1	0.0828	0.1260(0.1259)	0.0799(0.0799)	0.90(0.90)	0.79(0.79)
		0.2	0.2205	0.2657(0.2656)	0.1481(0.1484)	0.96(0.96)	0.90(0.90)
		0.3	0.5983	0.5276(0.5305)	0.1915(0.1919)	0.99(0.99)	0.96(0.95)
		0.4	0.9238	0.8491(0.8499)	0.1647(0.1645)	1.00(1.00)	1.00(1.00)
300	0.5	0.1	0.0828	0.1074(0.1079)	0.0446(0.0446)	0.91(0.91)	0.80(0.80)
		0.2	0.2205	0.2394(0.2392)	0.0821(0.0821)	0.97(0.97)	0.89(0.89)
		0.3	0.5983	0.5431(0.5422)	0.1617(0.1635)	0.99(0.99)	0.94(0.94)
		0.4	0.9238	0.8703(0.8704)	0.1394(0.1405)	1.00(1.00)	1.00(1.00)
500	0.5	0.1	0.0828	0.0984(0.0984)	0.0320(0.0320)	0.91(0.91)	0.86(0.87)
		0.2	0.2205	0.2287(0.2287)	0.0749(0.0751)	0.94(0.94)	0.89(0.89)
		0.3	0.5983	0.5185(0.5187)	0.1877(0.1888)	0.94(0.94)	0.88(0.88)
		0.4	0.9238	0.8343(0.8361)	0.1727(0.1725)	1.00(1.00)	1.00(1.00)

n	x_t(base)	x_s	True OR	$\widehat{OR}$	Stdev	95% coverage	90% coverage
100	0.2	0.4	0.4357	1.1889(1.2037)	0.9882(0.9857)	0.61(0.61)	0.53(0.53)
		0.6	7.0104	4.6104(4.6658)	4.5268(4.4634)	0.96(0.96)	0.91(0.91)
		0.8	0.3948	0.7666(0.7646)	0.6748(0.6723)	0.85(0.85)	0.74(0.73)
		1.0	0.0758	0.0595(0.0593)	0.1978(0.2050)	0.96(0.96)	0.88(0.87)
300	0.2	0.4	0.4357	0.6500(0.6461)	0.3658(0.3624)	0.88(0.88)	0.80(0.79)
		0.6	7.0104	5.6834(5.7262)	2.9581(2.9878)	0.98(0.98)	0.97(0.97)
		0.8	0.3948	0.5810(0.5917)	0.2791(0.2772)	0.87(0.87)	0.82(0.82)
		1.0	0.0758	0.0611(0.0638)	0.0688(0.0687)	0.92(0.91)	0.86(0.86)
500	0.2	0.4	0.4357	0.5266(0.5285)	0.1718(0.1725)	0.96(0.96)	0.93(0.93)
		0.6	7.0104	5.9574(5.9711)	2.5051(2.5162)	0.96(0.96)	0.92(0.92)
		0.8	0.3948	0.5170(0.5151)	0.2016(0.2024)	0.91(0.90)	0.86(0.85)
		1.0	0.0758	0.0540(0.0541)	0.0556(0.0559)	0.93(0.94)	0.86(0.86)

generate bootstrap samples from smoothing spline estimate. Then calculate the smoothing spline estimate of the parameter from these bootstrap samples. After calculating the estimate of the bias, the bias-corrected estimate is obtained. We re-evaluate the smoothing spline estimates of odds ratio and calculate the across-the-function coverage of Bayesian confidence intervals throughout simulations.

Simulated data were generated from a logistic distribution using the following test functions,

$$\eta_1(x) = \frac{1}{2}\beta_{10,5}(x) + \frac{1}{2}\beta_{7,7}(x) + \frac{1}{2}\beta_{5,10}(x) - 1$$

$$\eta_2(x) = 2\left(10^5 x^{11}(1-x)^6 + 10^3 x^3 (1-x)^{10}\right) - 2,$$

where $x = ((1:n) - .5)/n$ and $\beta_{a,b}(x) = [\Gamma(a+b)/\Gamma(a)\Gamma(b)]x^{a-1}(1-x)^{b-1}$. Note that $\eta_i, i = 1,2$, indicates the true function η. One hundred replicates were generated for each test function with samples of size $n = 100$, 300 and 500. For each replicates, the cubic smoothing splines η_λ was calculated with smoothing parameter selected by using AGACV score. To get odds ratios, we used $x_t = 0.5$ as the base and $x_s = 0.1$, $x_s = 0.2$, $x_s = 0.3$ and $x_s = 0.4$ for the test function η_1. For η_2, we used $x_t = 0.2$ as the base and $x_s = 0.4$, $x_s = 0.6$, $x_s = 0.8$ and $x_s = 1.0$. Note that he made an error to switch the test functions in his simulation results. Simulations were conducted to calculate the odds ratio estimates with 90% and 95% Bayesian confidence intervals with full basis($q = n$) and with reduced basis($q = 10n^{2/9}$) for 100 replicates.

Table 1 summarized the true odds ratios, estimated odds ratios as medians of 100 estimates of odds ratios, its standard deviations and the across-the-function coverage of 95% and 90% Bayesian confidence intervals with full basis. Those with reduced basis are inside the parenthesis. All results

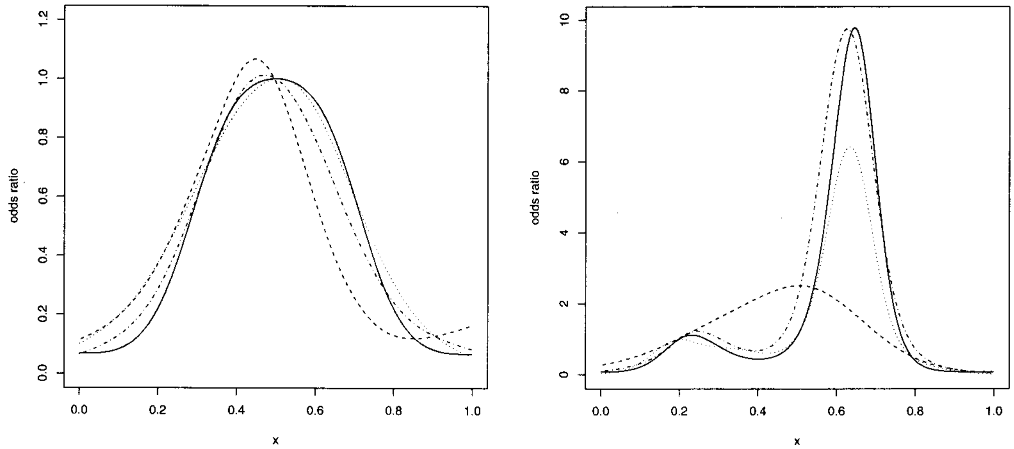

Figure 1: *Estimated odds ratio for* η_1 *(left) and* η_2 *(right) when n is increased: solid line is true odds ratio, dashed line is estimated odds ratio with* $n = 100$*, dotted line is estimated odds ratio with* $n = 300$*, and dotted-dashed line is estimated odds ratio with* $n = 500$*.*

showed no difference between with full basis and with reduced basis, but computations with reduced basis were much faster. Our estimates of odds ratio get closer to true odds ratios when the sample size gets larger, which is not surprising, yet they were biased toward unity as Wang (1997) pointed out. Bootstrap bias-corrected procedure wasn't working well here, a possible reason being that bootstrapping estimates were calculated from bootstrapping samples from the initial smoothing spline estimate whose performance was dependent on the smoothing parameter estimates. Thus, smoothing induced bias in estimates of η needs to be avoided while bootstrapping. In order to do it, more complex algorithm needs to follow. Instead of developing complex algorithm, taking larger sample help to solve such problems. Figure 1 showed the estimates of odds ratio when sample size gets increased. Figure 2 and 3 showed the pointwise coverage of 95% and 90% Bayesian confidence intervals for odds ratios from lower-dimensional approximations. The estimation of odds ratioes with η_1 showed relatively stable performance, whereas that with η_2 showed better performance when the sample size gets larger even though its across-the-function coverage in Table 1 is a bit decreasing which is not significant considering the corresponding confidence level. However, taking larger samples limits the practical use of smoothing splines. Lower-dimensional approximations is one solution to relieve the computation burden for estimation of odds ratio in smoothing splines settings. The proposed method can be used to analyze the binary data such as case-control data in medical studies.

6. Discussion

This paper discussed the odds ratio estimation in smoothing splines with the lower-dimensional approximation by means of simulations. Lowering the dimension of the estimating function space helped the computation of the smoothing spline estimate of the odds ratio more effectively for larger sample size while maintaining the performance of the estimate. Across-the-function coverage property of the pointwise Bayesian confidence intervals of smoothing spline estimate was evaluated. While the coverage shows the increasing pattern overall as the sample size increases compared to the corresponding confidence level, the performance of the estimate of odds ratio worked well.

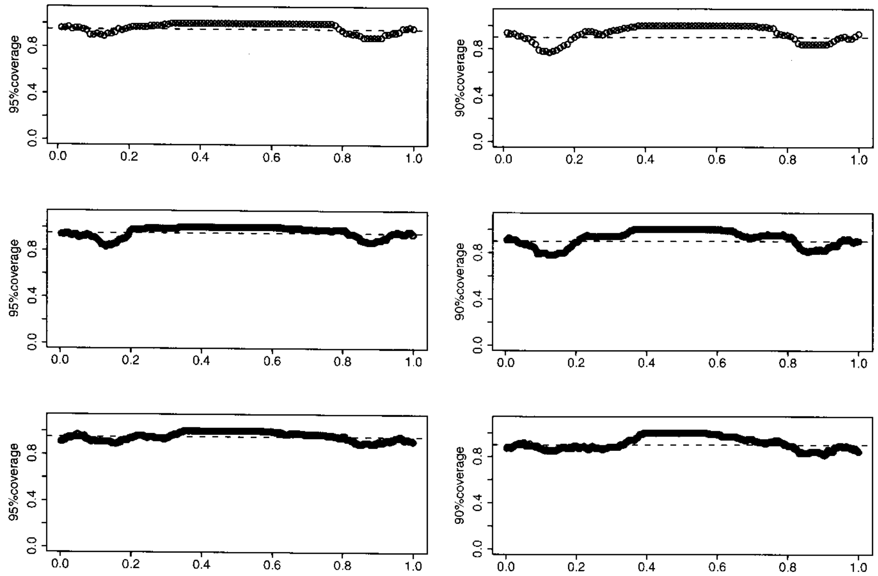

Figure 2: *Pointwise coverage of Bayesian confidence intervals of estimated odds ratio for for η_1. Left column is for 95% Bayesian confidence intervals and right column is for 90% Bayesian confidence intervals. From top to bottom, $n = 100, 300$ and 500.*

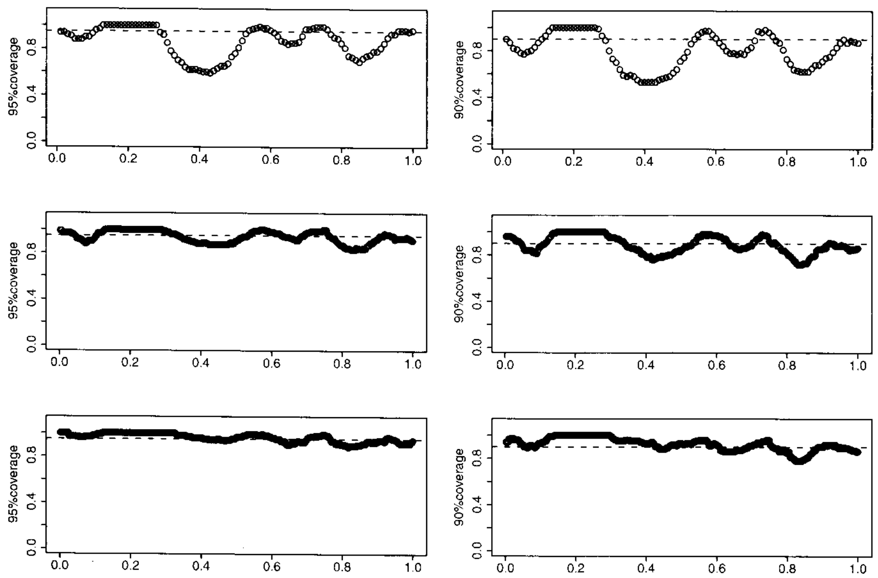

Figure 3: *Pointwise coverage of Bayesian confidence intervals of estimated odds ratio for for η_2. Left column is for 95% Bayesian confidence intervals and right column is for 90% Bayesian confidence intervals. From top to bottom, $n = 100, 300$ and 500.*

Acknowledgement

We thank the editor and referees for their constructive and detailed suggestions that led to significant improvement in the presentation of the paper.

Appendix A: Proof of Theorem 1

We adapt the proof of Gu (1992) and Section 5.4 in Gu (2002) and use the bayes model for the lower-dimensional approximations in Kim and Gu (2004). For $\eta = \eta_0 + \eta_1$, assume that $\eta_0(x) = \gamma^T \phi(x)$ with $\gamma \sim N(0, \tau^2 I)$ and η_1 has Gaussian prior with the covariance function

$$[\eta_1(x_i)\eta_1(x_2)] = bR_J\left(x_i, \mathbf{u}^T\right) Q^+ R_J(\mathbf{u}, x_j),$$

where Q^+ is the Moore-Penrose inverse of $Q = R_J(\mathbf{u}, \mathbf{u}^T)$. Let $\tau^2 \to \infty$. Then the likelihood of (η, γ) is proportional to

$$\exp\left\{-\frac{1}{2b}(\eta - S\gamma)^T \left(RQ^+R^T\right)^{-1} (\eta - S\gamma)\right\},$$

where RQ^+R^T has (i, j)th entry $R(x_i, \mathbf{u}^T)Q^+R(\mathbf{u}, x_j)$ and Q^+ is the Moore-Penrose inverse of $Q = R_J(\mathbf{u}, \mathbf{u}^T)$. Integrating out γ, the likelihood of η is

$$q(\eta) \propto \exp\left\{-\frac{1}{2b}\eta^T \left(N^{-1} - N^{-1}S\left(S^T N^{-1} S\right)^{-1} S^T N^{-1}\right)^{-1} \eta\right\},$$

where $N = RQ^+R^T$. Also the likelihood $p(Y|\eta)$ is proportional to

$$\exp\left\{-\frac{1}{\sigma^2}\sum_{i=1}^{n}\left\{Y_i\eta(x_i) - \log\left(1 + \exp(\eta(x_i))\right)\right\}\right\}.$$

Then, the posterior likelihood for η given $\mathbf{Y}$ is

$$p(\mathbf{Y}|\eta)q(\eta) \propto \exp\Bigg\{ -\frac{1}{\sigma^2}\sum_{i=1}^{n}\left\{Y_i\eta(\mathbf{x}_i) - \log(\mathbf{1} + \exp(\eta(\mathbf{x}_i)))\right\} - \frac{1}{2b}\eta^T \left(N^{-1} - N^{-1}S\left(S^T N^{-1} S\right)^{-1} S^T N^{-1}\right)^{-1} \eta\Bigg\}.$$

Now, we show that the posterior mode is equal to the solution to the penalized likelihood functional (2.1). The lower-dimensional approximating solution $\eta = S\mathbf{d} + R\mathbf{c}$ has $\mathbf{c}$ and $\mathbf{d}$ minimizing

$$-\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i(\boldsymbol{\phi}_i\mathbf{d} + \boldsymbol{\xi}_i\mathbf{c}) - \log(1 + \exp(\phi_i{}^*d^* + \xi_i c))\right\} + \frac{\lambda}{2}\mathbf{c}^T Q\mathbf{c}.$$

Taking derivatives with respect to $\mathbf{c}$ and $\mathbf{d}$ and setting them to zero, one has

$$S^T\mathbf{u} = 0,$$
$$R^T\mathbf{u} + n\lambda Q\mathbf{c} = 0,$$

where $\mathbf{v} = -\mathbf{Y} + \exp(\eta)/\{1 + \exp(\eta)\} = -\mathbf{Y} + p$. Differentiating $\log p(\mathbf{Y}|\eta)q(\eta)$ with respect to η and letting $n\lambda = \sigma^2/b$, one obtain

$$\begin{aligned}
&\mathbf{v} + n\lambda\left\{\left(N^{-1} - N^{-1}S\left(S^TN^{-1}S\right)^{-1}S^TN^{-1}\right)(S\mathbf{d} + R\mathbf{c})\right\} \\
&= \mathbf{v} + n\lambda\left\{\left(N^{-1}R\mathbf{c} - N^{-1}S\left(S^TN^{-1}S\right)^{-1}S^TN^{-1}R\mathbf{c}\right)\right\} = 0.
\end{aligned}$$

Therefore, the posterior mode of η given $\mathbf{Y}$ is equal to η_λ. Calculating the quadratic approximation of $p(\mathbf{Y}|\eta)$ at $\tilde{\eta}$, one has

$$p(\mathbf{Y}|\eta)q(\eta) \propto \tilde{p}(\mathbf{Y}|\eta)q(\eta),$$

where $\tilde{p}(\mathbf{Y}|\eta)$ is a quadratic approximation of $p(\mathbf{Y}|\eta)$ at $\tilde{\eta}$. In fact, $\tilde{p}(\mathbf{Y}|\eta)$ is Gaussian with zero mean and covariance σ^2W^{-1} for pseudo-observations $\tilde{\mathbf{Y}} = \tilde{\eta} - W^{-1}\tilde{v}$, where $\tilde{v} = -\mathbf{Y} + p$ and $W = \text{diag}(w_1, \ldots, w_n)$ with $w_i = p_i(1 + p_i)$. Then the Bayes model of Kim and Gu (2004) with pseudo-observations Y_w with variance σ^2W^{-1} and $M = RQ^+R^T + n\lambda W^{-1}$ can be applied directly to obtain the approximate posterior mean and variance from penalized weighted least square functional in Kim and Gu (2004) by replacing S by $W^{1/2}S$, R by $W^{1/2}R$, Y by $W^{1/2}Y$, and M by $M = RQ^+R^T + n\lambda W^{-1}$.

Appendix B: Proof of Theorem 2

To prove Theorem 2, it is enough to calculate the covariance of $\eta(\mathbf{x}_t)$ and $\eta(\mathbf{x}_s)$ given $\mathbf{Y}$. Following the proof of Theorem 3.1 in Gu and Wahba (1993), one can obtain the covariance of $\eta(\mathbf{x}_t)$ and $\eta(\mathbf{x}_s)$ in Gaussian case with the lower-dimensional approximation. Then the covariance of $\eta(\mathbf{x}_t)$ and $\eta(\mathbf{x}_s)$ in logistic framework is an approximation of that in weighted Gaussian framework. For $\mathbf{Y} = \eta + \epsilon$, where $E(\epsilon) = 0$, $E(\epsilon\eta) = 0$, $E(\eta\eta^T) = b\Sigma_{\eta\eta}$ and $E(\epsilon\epsilon^T) = \sigma^2W^{-1}$, assume that the random vectors g and h follow Gaussian distribution with zero mean and covariances $\Sigma_{gh^T} = b\Sigma_{gh}$, $\Sigma_{g\eta^T} = b\Sigma_{\eta g}$, $\Sigma_{\eta h^T} = b\Sigma_{\eta h}$. Then, the covariance between g and h given $\mathbf{Y}$ is $\text{cov}(g, h|\mathbf{Y}) = b(\Sigma_{gh} - \Sigma_{g\eta}(\Sigma_{\eta\eta} + n\lambda I)^{-1}\Sigma_{\eta h})$. Let $g = \eta(\mathbf{x}_t)$ and $h = \eta(\mathbf{x}_s)$. Then,

$$\begin{aligned}
\Sigma_{gh} &= E(\eta(\mathbf{x}_t)\eta(\mathbf{x}_s)) = b\left[\rho\boldsymbol{\phi}(\mathbf{x}_s)^T\boldsymbol{\phi}(\mathbf{x}_s) + R_J\left(\mathbf{x}_t, \mathbf{u}^T\right)Q^+R_J(\mathbf{u}, \mathbf{x}_s)\right] \\
\Sigma_{g\eta} &= E\left(\eta(\mathbf{x}_t)(\eta(x_1), \ldots, \eta(x_n))^T\right) = b\left[\rho\boldsymbol{\phi}(\mathbf{x}_s)^TS^T + R_J\left(\mathbf{x}_s, \mathbf{u}^T\right)Q^+(R_J(\mathbf{u}, x_1), \ldots, R_J(\mathbf{u}, x_n))^T\right] \\
\Sigma_{\eta h} &= b\left[\rho S\boldsymbol{\phi}(\mathbf{x}_t) + (R_J(x_1, \mathbf{u}), \ldots, R_J(x_n, \mathbf{u}))Q^+R_J(\mathbf{u}, \mathbf{x}_t)\right] \\
\Sigma_{\eta\eta} + n\lambda W^{-1} &= b\left[\rho SS^T + RQ^+R^T + n\lambda W^1\right] = b\left[\rho SS^T + M\right].
\end{aligned}$$

Thus, simple algebra gives

$$\begin{aligned}
\frac{\text{cov}[(\eta(\mathbf{x}_t), \eta(\mathbf{x}_s))|Y]}{b} &= \boldsymbol{\phi}(\mathbf{x}_s)^T\left[\rho I - \rho S^T\left(\rho SS^T + M\right)^{-1}\rho S\right]\boldsymbol{\phi}(\mathbf{x}_t) \\
&\quad - \rho\boldsymbol{\phi}(\mathbf{x}_s)^TS^T\left(\rho SS^T + M\right)^{-1}RQ^+R(\mathbf{t})^T \\
&\quad - R(\mathbf{s})Q^+R\left(\rho SS^T + M\right)^{-1}\rho S\boldsymbol{\phi}(\mathbf{x}_t) \\
&\quad + R(\mathbf{s})Q^+R(\mathbf{t})^T - R(\mathbf{s})Q^+R^T\left(\rho SS^T + M\right)^{-1}RQ^+R(\mathbf{t})^T,
\end{aligned}$$

where $R(\mathbf{s}) = (R_J(\mathbf{x}_s, u_1), \ldots, R_J(\mathbf{x}_s, u_n))^T$. Applying following formulas in Wahba (1983) produces the results.

$$\lim_{\rho\to\infty} \rho I - \rho S^T \left(\rho S S^T + M\right)^{-1} S\rho = \left(S^T M^{-1} S\right)^{-1}$$

$$\lim_{\rho\to\infty} \rho S^T \left(\rho S S^T + M\right)^{-1} = \left(S^T M^{-1} S\right)^{-1} S^T M^{-1}$$

$$\lim_{\rho\to\infty} \left(\rho S S^T + M\right)^{-1} = M^{-1} - M^{-1} S \left(S^T M^{-1} S\right)^{-1} S^T M^{-1}.$$

References

Gu, C. (1992). Penalized likelihood regression: a Bayesian analysis, *Statistica Sinica,* **2**, 255–264.

Gu, C. (2002). *Smoothing Spline ANOVA models,* Springer-Verlag.

Gu, C. and Kim, Y.-J. (2002). Penalized likelihood regression: General formulation and efficient approximation, *Canadian Journal of Statistics,* **30**, 619–628.

Gu, C. and Wahba, G. (1993). Smoothing spline ANOVA with component-wise Bayesian confidence intervals, *Journal of computational and graphical statistics,* **2**, 97–117.

Gu, C. and Xiang, D. (2001). Cross-validating non-Gaussian data: Generalized approxmate cross-validation revisited, *Journal of computational and graphical statistics,* **10**, 581–591.

Kim, Y.-J. (2003). Smoothing spline regression: scalable computation and cross-validation, *Ph.D. diss.*, Purdue University.

Kim, I., Cohen, N. D. and Carroll, R. J. (2003). Semiparametric regression splines in matched case-control studies, *Biometrics,* **59**, 1158–1169.

Kim, Y.-J. and Gu, C. (2004). Smoothing spline Gaussian regression: More scalable computation via efficient approximation, *Journal of the Royal Statistical Society Series B,* **66**, 337–356.

Lin, X., G. Wabha, D. Xiang, F. Gao, R. Klein, and Klein, B. E. K. (2000). Smoothing spline ANOVA models for large data sets with Bernoulli observations and the randomized GACV, *The Annals of Statistics,* **28**, 1570–1600.

Luo, Z. and Wahba, G. (1997). Hybrid adaptive splines, *Journal of the American Statistical Association,* **92**, 107–116.

Wang, Y. (1997). Odds ratio estimation in bernoulli smoothing spline ANOVA models, *Statistician,* **48**, 49–56.

Wahba, G. (1983). Bayesian "confidence interval" for the cross-validated smoothing spline, *Journal of the Royal Statistical Society Series B,* **45**, 133–150.

Wahba, G., Wang, Y., Gu, C., Klein, R. and Klein, B. (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological study of diabetic retinopathy, *Annals of Statistics,* **23**, 1865–1895.

Xiang, D. and Wahba, G. (1998). Approximate smoothing spline methods for large data sets in the binary case, *Proceedings of the 1997 ASA Joint Statistical Meetings, Biometrics Section,* 94–98.

Received April 2009; Accepted June 2009